

# Induced subarrays of Latin squares without repeated symbols

R. Julian R. Abel

School of Mathematics and Statistics  
University of New South Wales  
N.S.W. 2052  
Australia

[r.j.abel@unsw.edu.au](mailto:r.j.abel@unsw.edu.au)

Nicholas J. Cavenagh

Department of Mathematics  
The University of Waikato  
Private Bag 3105  
Hamilton 3240, New Zealand

[nickc@waikato.ac.nz](mailto:nickc@waikato.ac.nz)

Jaromy Kuhl

Department of Mathematics and Statistics  
University of West Florida  
Pensacola Florida, 32514

[jkuhl@uwf.edu](mailto:jkuhl@uwf.edu)

Submitted: May 21, 2012; Accepted: Feb 24, 2013; Published: Mar 1, 2013

## Abstract

We show that for any Latin square  $L$  of order  $2m$ , we can partition the rows and columns of  $L$  into pairs so that at most  $(m + 3)/2$  of the  $2 \times 2$  subarrays induced contain a repeated symbol. We conjecture that any Latin square of order  $2m$  (where  $m \geq 2$ , with exactly five transposition class exceptions of order 6) has such a partition so that *every*  $2 \times 2$  subarray induced contains no repeated symbol. We verify this conjecture by computer when  $m \leq 4$ .

*Keywords:* Latin square, 2-partition, conjugate, isotopic, transposition class, discrepancy, potential.

*AMS subject classification:* 05B15

## 1 Introduction

For each integer  $n$ , we define  $[n]$  to be the set of integers from 1 to  $n$  inclusive. If  $S = \{s_1, s_2, \dots, s_n\}$  is a set of size  $n$ , then a Latin square of order  $n$  (over  $S$ ) is a  $n \times n$  array such that each element (or *symbol*) of  $S$  occurs precisely once in each row and once in each column. If we label the rows and columns with sets  $R = \{r_1, r_2, \dots, r_n\}$  and  $C = \{c_1, c_2, \dots, c_n\}$  respectively, we may think of a Latin square as a set of ordered triples  $(r_i, c_j, s_k)$ , where symbol  $s_k$  occurs at the intersection of row  $r_i$  and column  $c_j$ . We may

define a binary operation  $\circ$  on the sets  $R$  and  $C$  of a Latin square  $L$  such that  $r_i \circ c_j = s_k$  if and only if  $(r_i, c_j, s_k) \in L$ .

We say that a partition  $\mathcal{P}$  of a set  $X$  is a  $k$ -partition if  $|P| = k$  for each  $P \in \mathcal{P}$ . Henceforth  $n$  is even and  $m = n/2$ . Let  $\mathcal{R} = \{R_1, R_2, \dots, R_m\}$  and  $\mathcal{C} = \{C_1, C_2, \dots, C_m\}$  be 2-partitions of the sets  $R$  and  $C$  respectively. For each  $R' \in \mathcal{R}$  and  $C' \in \mathcal{C}$ , we define  $S(R', C')$  to be the  $2 \times 2$  subarray of  $L$  induced by  $R'$  and  $C'$ ; that is:

$$S(R', C') := \{(r_i, c_j, r_i \circ c_j) \mid r_i \in R', c_j \in C'\}.$$

We say that the pair of 2-partitions  $(\mathcal{R}, \mathcal{C})$  is *good* if for each  $R' \in \mathcal{R}$  and  $C' \in \mathcal{C}$ ,  $S(R', C')$  contains no repeated symbols (i.e. four distinct symbols). Two Latin squares are said to be in the same *isotopy class* if one can be obtained from the other by rearranging rows, rearranging columns and/or rearranging symbols. Two Latin squares are in the same *transposition class* if one can be obtained from the other by isotopy and/or taking the transpose. It is clear that if a Latin square has a good pair of 2-partitions, then so does every Latin square in that transposition class. If  $\phi$  is any one of the six permutations defined on the set  $\{1, 2, 3\}$ , then the  $(\phi(1), \phi(2), \phi(3))$ -conjugate of a Latin square  $L$  is formed by replacing each triple  $(t(1), t(2), t(3))$  associated with  $L$  by the triple  $(t(\phi(1)), t(\phi(2)), t(\phi(3)))$ . Latin squares which are related by conjugate and/or isotopy are said to lie in the same *main class*. However not all conjugacies preserve 2-partitions, for instance later in Table 1 we will see three squares of order 6 that have no good pair of 2-partitions, even though some of their conjugates in Table 4 do. Thus transposition class is the broadest class for the purposes of this paper.

We now present the following conjecture.

**Conjecture 1** *Let  $L$  be a Latin square of even order  $n$ , where  $n \geq 4$ . Then a good pair of 2-partitions of the rows and columns of  $L$  exists, with five transposition class exceptions (the five Latin squares of order 6 in Table 1).*

The first, third, fourth and fifth squares in Table 1 are all symmetric (and are respectively the third, fourth, fifth, and sixth squares in [2], p137). The second square is the  $(2, 3, 1)$ -conjugate of the first square. For all these squares, we give a pair of 2-partitions  $(\mathcal{R}, \mathcal{C})$  with the smallest possible total number of repeated symbols in the induced  $2 \times 2$  squares. The number of such repeats is two for the second square, and one for the others.

The above conjecture is supported heuristically for small values of  $n$ .

**Theorem 2** *Let  $L$  be a Latin square of even order  $n$ , with  $4 \leq n \leq 8$ . Then a good pair of 2-partitions of the rows and columns of  $L$  exists, with five transposition class exceptions (the five Latin squares of order 6 in Table 1).*

*Proof.* We begin with  $n = 4$  (see [2] for definitions). Representatives from the two transposition classes of Latin squares of order 4 are given in Table 2. For these, set  $\mathcal{R} = \{\{r_1, r_2\}, \{r_3, r_4\}\}$  and  $\mathcal{C} = \{\{c_1, c_3\}, \{c_2, c_4\}\}$ . Then the pair of 2-partitions  $(\mathcal{R}, \mathcal{C})$  is good.

Table 1: Representatives for the five transposition classes of Latin squares of order 6 with no good pair of 2-partitions.

1	2	3	4	5	6
2	1	4	5	6	3
3	4	1	6	2	5
4	5	6	1	3	2
5	6	2	3	1	4
6	3	5	2	4	1

$\mathcal{R} : (r_1, r_4), (r_2, r_5), (r_3, r_6)$   
 $\mathcal{C} : (c_1, c_6), (c_2, c_3), (c_4, c_5)$

1	2	3	4	5	6
2	1	6	3	4	5
3	5	1	2	6	4
4	6	5	1	2	3
5	3	4	6	1	2
6	4	2	5	3	1

$\mathcal{R} : (r_1, r_4), (r_2, r_5), (r_3, r_6)$   
 $\mathcal{C} : (c_2, c_3), (c_4, c_6), (c_1, c_5)$

1	2	3	4	5	6
2	1	4	5	6	3
3	4	1	6	2	5
4	5	6	1	3	2
5	6	2	3	4	1
6	3	5	2	1	4

$\mathcal{R} : (r_1, r_6), (r_2, r_4), (r_3, r_5)$   
 $\mathcal{C} : (c_1, c_2), (c_4, c_5), (c_3, c_6)$

1	2	3	4	5	6
2	1	4	5	6	3
3	4	2	6	1	5
4	5	6	2	3	1
5	6	1	3	4	2
6	3	5	1	2	4

$\mathcal{R} : (r_1, r_6), (r_2, r_4), (r_3, r_5)$   
 $\mathcal{C} : (c_1, c_2), (c_4, c_5), (c_3, c_6)$

1	2	3	4	5	6
2	1	4	5	6	3
3	4	5	6	1	2
4	5	6	3	2	1
5	6	1	2	3	4
6	3	2	1	4	5

$\mathcal{R} : (r_1, r_5), (r_2, r_4), (r_3, r_6)$   
 $\mathcal{C} : (c_1, c_6), (c_2, c_3), (c_4, c_5)$

Table 2: Representatives for the two transposition classes of Latin squares of order 4.

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

1	2	3	4
2	1	4	3
3	4	2	1
4	3	1	2

There are 1,676,267 isotopy classes of Latin squares of order 8 [2, 3]; a representative for each of these classes is given at [3]. We ran a program which reads these squares in batches of 100,000 and can produce a good pair of 2-partitions for all squares in each batch in about 3 minutes. A good pair of 2-partitions was found for all representatives;

in fact, even if we add an extra restriction that rows 1 to 4 all lie in different pairs for the 2-partitions, then for almost all squares of order 8, a good pair of 2-partitions can be found. Table 3 gives one square with a good pair of 2-partitions, but none satisfying this extra restriction.

Table 3: A Latin square of order 8 and a good pair of 2-partitions for it.

1	2	3	4	5	6	7	8
2	1	7	8	3	4	6	5
3	6	8	2	7	5	1	4
4	5	2	7	6	8	3	1
5	8	6	3	2	1	4	7
6	7	4	5	1	2	8	3
7	4	1	6	8	3	5	2
8	3	5	1	4	7	2	6

$$\mathcal{R} : (r_1, r_4), (r_2, r_6), (r_3, r_8), (r_5, r_7)$$

$$\mathcal{C} : (c_1, c_5), (c_2, c_4), (c_3, c_8), (c_6, c_7)$$

There are twelve main classes of Latin squares of order 6 [2]. In seven of these, each square is isotopic to all its conjugates (and hence also to every other square in the same main class). Within each of the other five main classes, there are three isotopic classes and two transposition classes. Thus there is a total of  $7 + (5 \times 3) = 22$  isotopic classes of Latin squares of order 6 [2, 3] and  $7 + (5 \times 2) = 17$  transposition classes. If  $S$  is any square in one of these last five main classes that is isotopic to its transpose, then  $S$ , together with its  $(2, 3, 1)$  and  $(3, 1, 2)$ -conjugates are representatives for the three isotopic classes in the same main class as  $S$ . Also,  $S$  and its  $(2, 3, 1)$ -conjugate are representatives for the two transposition classes in this main class.

For order 6, representatives for the twelve transposition classes not given in Table 1 (together with a pair of good 2-partitions for each one) are given in Tables 4 and 5. The eight squares in Table 5 are all isotopic to their transposes; in fact, all except the fifth are isotopic to all their conjugates [2]. The four squares in Table 4 are not isotopic to their transposes; these squares are the  $(2,3,1)$ -conjugates of respectively, the third, fourth and fifth squares in Table 1 and the fifth square in Table 5.  $\square$

## 2 The problem with a particular approach

Before we prove our main result (Theorem 9), it is worth demonstrating why a particular method for solving Conjecture 1 cannot work. As a first approach to the problem, one might assume a fixed 2-partition of the rows, then try adjusting the 2-partition of the columns. In this section we show that such an approach cannot solve Conjecture 1 directly. In particular, the results in this section imply the following theorem.

Table 4: Representatives and a good pair of 2-partitions for four transposition classes of Latin squares of order 6.

1	2	3	4	5	6
2	1	6	3	4	5
3	5	1	2	6	4
4	6	5	1	2	3
6	3	4	5	1	2
5	4	2	6	3	1

$\mathcal{R} : (r_1, r_5), (r_2, r_6), (r_3, r_4)$   
 $\mathcal{C} : (c_1, c_2), (c_3, c_5), (c_4, c_6)$

1	2	3	4	5	6
2	1	6	3	4	5
5	3	1	2	6	4
6	4	5	1	2	3
3	6	4	5	1	2
4	5	2	6	3	1

$\mathcal{R} : (r_1, r_5), (r_2, r_4), (r_3, r_6)$   
 $\mathcal{C} : (c_1, c_4), (c_2, c_3), (c_5, c_6)$

1	2	3	4	5	6
2	1	6	3	4	5
5	6	1	2	3	4
6	5	4	1	2	3
3	4	5	6	1	2
4	3	2	5	6	1

$\mathcal{R} : (r_1, r_4), (r_2, r_5), (r_3, r_6)$   
 $\mathcal{C} : (c_1, c_3), (c_2, c_6), (c_4, c_5)$

1	2	3	6	5	4
2	1	5	3	4	6
3	4	1	5	6	2
4	3	6	1	2	5
5	6	2	4	1	3
6	5	4	2	3	1

$\mathcal{R} : (r_1, r_5), (r_2, r_4), (r_3, r_6)$   
 $\mathcal{C} : (c_1, c_4), (c_2, c_6), (c_3, c_5)$

**Theorem 3** For each  $m \geq 3$ , there exist a Latin square of order  $2m$  and a 2-partition  $\mathcal{R}$  of its rows such that for each 2-partition  $\mathcal{C}$  of the columns, the pair  $(\mathcal{R}, \mathcal{C})$  is bad.

We give a construction to show that such a Latin square, along with a 2-partition  $\mathcal{R}$  of the rows, exists for every even order  $2m \geq 6$ . Without loss of generality, we may assume that for  $\mathcal{R} = \{R_1, \dots, R_m\}$ ,  $R_i = \{2i - 1, 2i\}$  for each  $i \in [m]$ . Theorem 3 is true if we can construct a Latin square which contains the set of triples

$$\{(1, 1, 1), (1, 2, 2), (2, 1, 2), (2, 2, 1)\} \cup \{(2i - 1, 1, 2i - 1), (2i, 1, 2i), (2i - 1, 2i - 1, 2i), (2i, 2i, 2i - 1) : 2 \leq i \leq m\}.$$

Below is an example of the triples as a partial Latin square of order 6 ( $m = 3$ ).

1	2				
2	1				
3		4			
4			3		
5				6	
6					5

Note that  $S(R_1, \{c_1, c_2\})$  contains symbols 1 and 2 twice. Furthermore,  $S(R_i, \{c_1, c_{2i}\})$  contains symbol  $2i - 1$  twice and  $S(R_i, \{c_1, c_{2i-1}\})$  contains symbol  $2i$  twice for each  $i$ ,

Table 5: Representatives and a good pair of 2-partitions for eight transposition classes of Latin squares of order 6.

1	2	3	4	5	6
2	1	4	3	6	5
3	4	5	6	1	2
4	3	6	5	2	1
5	6	1	2	3	4
6	5	2	1	4	3

$$\mathcal{R} : (r_1, r_4), (r_2, r_5), (r_3, r_6)$$

$$\mathcal{C} : (c_1, c_5), (c_2, c_6), (c_3, c_4)$$

1	2	3	4	5	6
2	1	4	3	6	5
3	4	5	6	1	2
4	3	6	5	2	1
5	6	1	2	4	3
6	5	2	1	3	4

$$\mathcal{R} : (r_1, r_4), (r_2, r_5), (r_3, r_6)$$

$$\mathcal{C} : (c_1, c_3), (c_2, c_4), (c_5, c_6)$$

1	2	3	4	5	6
2	1	4	3	6	5
3	5	1	6	2	4
4	6	2	5	1	3
5	3	6	1	4	2
6	4	5	2	3	1

$$\mathcal{R} : (r_1, r_4), (r_2, r_5), (r_3, r_6)$$

$$\mathcal{C} : (c_1, c_3), (c_2, c_5), (c_4, c_6)$$

1	2	3	4	5	6
2	1	4	3	6	5
3	5	1	6	2	4
4	6	2	5	1	3
5	3	6	2	4	1
6	4	5	1	3	2

$$\mathcal{R} : (r_1, r_4), (r_2, r_5), (r_3, r_6)$$

$$\mathcal{C} : (c_1, c_3), (c_2, c_5), (c_4, c_6)$$

1	2	3	4	5	6
2	1	6	5	3	4
3	4	1	2	6	5
6	5	2	1	4	3
5	3	4	6	1	2
4	6	5	3	2	1

$$\mathcal{R} : (r_1, r_4), (r_2, r_5), (r_3, r_6)$$

$$\mathcal{C} : (c_1, c_3), (c_2, c_4), (c_5, c_6)$$

1	2	3	4	5	6
2	1	4	3	6	5
3	5	1	6	4	2
4	6	5	1	2	3
5	3	6	2	1	4
6	4	2	5	3	1

$$\mathcal{R} : (r_1, r_4), (r_2, r_5), (r_3, r_6)$$

$$\mathcal{C} : (c_1, c_2), (c_3, c_4), (c_5, c_6)$$

1	2	3	4	5	6
2	1	4	5	6	3
3	4	2	6	1	5
4	6	5	2	3	1
5	3	6	1	2	4
6	5	1	3	4	2

$$\mathcal{R} : (r_1, r_4), (r_2, r_5), (r_3, r_6)$$

$$\mathcal{C} : (c_1, c_2), (c_3, c_4), (c_5, c_6)$$

1	2	3	4	5	6
2	3	1	5	6	4
3	1	2	6	4	5
4	6	5	2	1	3
5	4	6	3	2	1
6	5	4	1	3	2

$$\mathcal{R} : (r_1, r_4), (r_2, r_5), (r_3, r_6)$$

$$\mathcal{C} : (c_1, c_6), (c_2, c_5), (c_3, c_4)$$

$2 \leq i \leq m$ . It follows that for each  $j \in [2m]$ , column  $c_1$  cannot be paired with column  $c_j$  without creating an induced subarray with repeated symbols. Thus, if there is a Latin square of order  $2m$  containing the triples defined above, then  $(\mathcal{R}, \mathcal{C})$  is bad for each 2-partition  $\mathcal{C}$  of columns.

**Lemma 4** *If  $m \geq 3$ , then there exists a Latin square  $L$  of order  $2m$  containing the above partial Latin square.*

*Proof.* Let  $L_1(\circ)$  be the Latin square of order  $m$  such that  $i \circ j = i + j - 1 \pmod{m}$  for each  $i, j \in [m]$ . Let  $L_2(\star)$  be an idempotent Latin square of order  $m$  (i.e.  $i \star i = i$ , for each  $i \in [m]$ ). These are known to exist for all orders  $m \geq 3$  (see [2]).

Next, we create a Latin square of order  $2m$  by positioning a  $2 \times 2$  subarray at each cell  $(i, j) \in [m] \times [m]$  of an  $m \times m$  grid. Firstly, for each cell  $(i, j)$  for which  $i \circ j = i \star j$  (this includes the case  $(i, j) = (1, 1)$ ) we place the  $2 \times 2$  subarray

$2(i \circ j) - 1$	$2(i \circ j)$
$2(i \circ j)$	$2(i \circ j) - 1$

To determine what  $2 \times 2$  array should be placed in each of the other cells, we first have to construct for each  $y \in [m]$ , a graph  $G_y$  whose vertices are the cells  $(i, j) \in [m] \times [m]$  such that  $i \circ j = y$  or  $i \star j = y$  (but not both). Two distinct vertices  $(i_1, j_1)$  and  $(i_2, j_2)$  in  $G_y$  are said to be adjacent if

1.  $i_1 = i_2$  or  $j_1 = j_2$ , and
2.  $i_1 \circ j_1 = i_2 \star j_2 = y$  or  $i_1 \star j_1 = i_2 \circ j_2 = y$ .

With adjacency defined this way, observe that  $G_y$  is a union of vertex-disjoint cycles of even length. Thus every vertex in  $G_y$  has two vertices adjacent to it.

For each  $y \in [m]$ ,  $y \neq 1$ , there is one cycle in  $G_y$  containing two cells in column 1. Let cell  $(y, 1)$  be the first cell in this cycle; subsequent cells will be chosen so that they are connected to the previous one. In the  $2 \times 2$  array placed in the first cell, place  $2y - 1$  in the top left entry and  $2y$  in the bottom left entry.

Let cell  $(y, y)$  be the second cell of this cycle. In the  $2 \times 2$  array placed in this second cell (and every remaining cell of the cycle except the last) place  $2y$  and  $2y - 1$  in two vacant, diagonally opposite squares of the  $2 \times 2$  subarray in that cell, so that  $2y$  and  $2y - 1$  are not in the same row or column of our  $2m \times 2m$  array  $L$  as they were in the previous  $2 \times 2$  subarray. In addition, for the  $2 \times 2$  subarray in the second cell of the cycle,  $2y$  must be in the top left entry, and  $2y - 1$  must be in the bottom right entry.

For the last cell of this cycle (this cell will appear in column 1) we place  $2y$  and  $2y - 1$  in the two entries in the right hand column of the  $2 \times 2$  subarray in this cell. Again, this should be done so that  $2y$  and  $2y - 1$  are not in the same row of our  $2m \times 2m$  array  $L$  as they were in the  $2 \times 2$  array placed in the previous cell.

Finally, for each  $y \in [m]$ , consider each cycle in  $G_y$  not containing any cells  $(i, j)$  with  $j = 1$ . If  $x$  is the length of such a cycle, then as mentioned earlier,  $x$  is always even; also, we can partition the vertices of this cycle into two independent sets of size  $x/2$  (i.e. no two vertices (cells) in either set are adjacent). For each  $2 \times 2$  array placed in a cell of the first set, we place  $2y$  in two vacant, diagonally opposite entries of that  $2 \times 2$  array. For

each  $2 \times 2$  array placed in a cell of the other set, we place  $2y - 1$  in two vacant diagonally opposite entries.

This completes the construction. □

Below is an example of the previous lemma with  $m = 4$ :

1	2	3	4
2	3	4	1
3	4	1	2
4	1	2	3

$L_1(\circ)$

1	4	2	3
3	2	4	1
4	1	3	2
2	3	1	4

$L_2(\star)$

1	2	7	4	3	6	5	8
2	1	3	8	5	4	7	6
3	5	4	6	7	8	1	2
4	6	5	3	8	7	2	1
5	7	8	1	6	2	3	4
6	8	1	7	2	5	4	3
7	4	6	2	1	3	8	5
8	3	2	5	4	1	6	7

### 3 Proving the main result

If a pair of 2-partitions  $(\mathcal{R}, \mathcal{C})$  for a Latin square is not *good*, then we say it is *bad*. Informally, in what follows we wish to take any even order Latin square and find a 2-partition with minimum “badness”. A  $2 \times 2$  subarray may of course contain *two* pairs of repeated symbols. It is helpful then to call the existence of a pair of repeated symbols within an induced subarray a *discrepancy*. Thus an induced subarray may have either 0, 1 or 2 discrepancies. We define  $\Delta(\mathcal{R}, \mathcal{C})$  to be the total number of discrepancies in all of the induced subarrays. We will focus on minimizing the function  $\Delta$ .

For each pair of rows  $R' \in \mathcal{R}$ , we define  $\Delta(R')$  (or  $\Delta_{\mathcal{C}}(R')$  when ambiguity may arise) to be the number of discrepancies *within* the rows of  $R'$ . We define  $\Delta(C')$  similarly for each  $C' \in \mathcal{C}$ . Observe that

$$\Delta(\mathcal{R}, \mathcal{C}) = \sum_{i=1}^m \Delta(R_i) = \sum_{i=1}^m \Delta(C_i).$$

Our general approach is to reduce  $\Delta$  by making minimal adjustments to the sets  $\mathcal{R}$  or  $\mathcal{C}$ , respectively. Formally, suppose that  $\mathcal{C}$  is a 2-partition with  $C_i = \{c_1, c_2\}$  and  $C_j = \{c_3, c_4\}$ . Then there are precisely two options for changing  $\mathcal{C}$  by altering  $\{C_i, C_j\}$  but keeping each other pair in  $\mathcal{C}$  fixed. We can substitute  $\{C_i, C_j\}$  with either  $\{\{c_1, c_3\}, \{c_2, c_4\}\}$  or  $\{\{c_1, c_4\}, \{c_2, c_3\}\}$ . We call such a substitution a *trade* on  $C_i$  and  $C_j$ . We define a trade between two elements of  $\mathcal{R}$  in the same way.

Each trade may destroy certain discrepancies but create others. Thus we introduce the idea of a discrepancy being potentially created by a trade. Suppose there exists a symbol  $s \in S$ , a pair of rows  $R \in \mathcal{R}$  and distinct pairs of columns  $C, C' \in \mathcal{C}$  such that  $s$  occurs in the subarrays  $S(R, C)$  and  $S(R, C')$ . We say that the 4-tuple  $(s, R, C, C')$  is a *potential* (or *column potential*). (We may define *row potentials* similarly in the transpose.)

It is important to note that a potential leads to a discrepancy via precisely one of the two available trades. It is thus useful to label the two choices of trade as types 1 and 2



for convenience, in some arbitrary way. We define  $d^+(C_i \leftrightarrow C_j, t)$  and  $d^-(C_i \leftrightarrow C_j, t)$  to be the number of discrepancies created and destroyed (respectively) after a trade of type  $t \in \{1, 2\}$  between pairs of columns  $C_i$  and  $C_j$ . If  $d^+(C_i \leftrightarrow C_j, t) < d^-(C_i \leftrightarrow C_j, t)$ , then  $\Delta(\mathcal{R}, \mathcal{C}') < \Delta(\mathcal{R}, \mathcal{C})$ , where  $\mathcal{C}'$  is the new column partition after the trade. This is our chief strategy in reducing  $\Delta$ . In some cases, our trades will initially increase the number of discrepancies; however, we will always ensure that overall, a reduction in the number of discrepancies results.

Before we begin our proofs, the following observations are also useful. Suppose that  $(s, R, C, C')$  and  $(s', R', C, C')$  are distinct potentials. Then it is possible that either  $s = s'$  or  $R = R'$ ; but we cannot have  $(s, R) = (s', R')$ . Indeed, a potential  $(s, R, C, C')$  is completely identified by the pair  $(s, R)$ , since  $s$  occurs once in each of the induced subarrays  $S(R, C)$  and  $S(R, C')$ . Also if  $s \in S$  and  $C \in \mathcal{C}$  are fixed, then if there is no discrepancy involving both  $s$  and  $C$ , there are two column potentials of the form  $(s, R', C, C')$  with  $R' \in \mathcal{R}$  and  $C' \in \mathcal{C}$ . Since any symbol and pair of columns are both involved in either one discrepancy or two column potentials, the following lemma is immediate.

**Lemma 5** *The total number of column potentials  $(s, R', C, C')$  involving a given pair of columns  $C \in \mathcal{C}$  is equal to  $4m - 2\Delta(C)$ .*

**Lemma 6** *Let  $(\mathcal{R}, \mathcal{C})$  be a pair of 2-partitions of the rows and columns of a Latin square  $L$ . Suppose that  $C_1$  and  $C_2$  are distinct elements of  $\mathcal{C}$  and that  $\Delta(C_1) = k$  and  $\Delta(C_2) = l$ , for integers  $k, l \geq 0$ . If either  $(k \geq 2$  and  $l \geq 1)$  or  $k \geq 3$ , then there exists a 2-partition  $\mathcal{C}'$  (obtained by a single trade including  $C_1$ ) such that  $\Delta(\mathcal{R}, \mathcal{C}') \leq \Delta(\mathcal{R}, \mathcal{C}) - k + 1$ .*

*Proof.* Suppose first that  $k \geq 2$  and  $l \geq 1$ . We wish to make a trade on  $C_1$  and  $C_j$ ,  $j \neq 1$ , of type  $t \in \{1, 2\}$  so that

$$d^-(C_1 \leftrightarrow C_j, t) - d^+(C_1 \leftrightarrow C_j, t) > k - 2.$$

For the sake of contradiction, suppose that no such trade exists. Then for each such trade,

$$d^-(C_1 \leftrightarrow C_j, t) - d^+(C_1 \leftrightarrow C_j, t) \leq k - 2. \tag{1}$$

Suppose first that  $j = 2$ . Let  $t \in \{1, 2\}$ . From two assumptions of the lemma ( $\Delta(C_1) = k$  and  $\Delta(C_2) \geq 1$ ), we have  $d^-(C_1 \leftrightarrow C_2, t) \geq k + 1$ . Hence from (1),  $d^+(C_1 \leftrightarrow C_2, t) \geq 3$  for each  $t \in \{1, 2\}$ . Since there are two choices for  $t$ , there must exist at least six pairwise distinct potentials involving  $C_1$  and  $C_2$ .

Next consider the case  $j > 2$ . Let  $t \in \{1, 2\}$ . Then  $d^-(C_i \leftrightarrow C_j, t) \geq k$ ; thus (1) yields:  $d^+(C_1 \leftrightarrow C_j, t) \geq 2$ . Since there are two choices for  $t$ , there must exist at least four pairwise distinct potentials involving  $C_1$  and  $C_j$ . Thus in total there are at least  $6 + 4(m - 2)$  pairwise distinct (column) potentials involving  $C_1$ . Applying Lemma 5,  $6 + 4(m - 2) \leq 4m - 2k$ , which contradicts  $k \geq 2$ .

Now suppose that  $k \geq 3$ . By similar reasoning to the above, for each  $j \geq 2$  and  $t \in \{1, 2\}$ ,  $d^+(C_1 \leftrightarrow C_j, t) \geq 2$ . Therefore, there are at least  $4(m - 1)$  pairwise distinct (column) potentials involving the pair of columns  $C_1$ . Applying Lemma 5,  $4(m - 1) \leq 4m - 2k$ , which contradicts  $k \geq 3$ .  $\square$

Using Lemma 6 and its transpose analogue repeatedly, we have the following corollary to Lemma 6.

**Corollary 7** *Let  $L$  be a Latin square of order  $2m$ . Suppose there exists a pair of 2-partitions  $(\mathcal{R}, \mathcal{C})$  of  $L$  such that  $\Delta(\mathcal{R}, \mathcal{C}) = \beta$ . Then there is a pair of 2-partitions  $(\mathcal{R}', \mathcal{C}')$  of  $L$  such that  $\Delta(\mathcal{R}', \mathcal{C}') \leq \beta$  and either:*

1.  $\Delta(R), \Delta(C) \in \{0, 1\}$  for each  $R \in \mathcal{R}'$  and  $C \in \mathcal{C}'$ ; or
2.  $\Delta(\mathcal{R}', \mathcal{C}') = 2$ .

*Proof.* If there are elements  $C_i, C_j \in \mathcal{C}$  such that  $\Delta(C_i) \geq 2$  and  $\Delta(C_j) \geq 1$ , then by Lemma 6 there is a 2-partition of the columns of  $L$  reducing the current number of discrepancies in  $L$ . Similarly, the function  $\Delta$  is reduced if there are elements  $R_i, R_j \in \mathcal{R}$  where  $\Delta(R_i) \geq 2$  and  $\Delta(R_j) \geq 1$ . Furthermore, if there are no such  $i, j$ , but there exists  $i$  such that  $\Delta(C_i) \geq 3$  (or  $\Delta(R_i) \geq 3$ ), then  $\Delta$  is reduced. Thus repeatedly invoking Lemma 6 and its transpose analogue continues to lower  $\Delta$  until there is a pair of 2-partitions  $(\mathcal{R}', \mathcal{C}')$  satisfying the conditions of this lemma.  $\square$

**Lemma 8** *Let  $(\mathcal{R}, \mathcal{C})$  be a pair of 2-partitions of the rows and columns of a Latin square  $L$ . Let  $\mathcal{R} = \{R_1, R_2, \dots, R_m\}$  and let  $\Delta(R_i) = k_i$  for each  $i \in [m]$ . Suppose furthermore that  $\alpha$  is an integer in  $[m]$  such that either  $(\alpha \geq 2, k_\alpha = 2$  and  $k_t \geq 1$  for some  $t \in [m], t \neq \alpha)$  or  $k_\alpha \geq 3$ . Then there exists a 2-partition  $\mathcal{R}'$  such that  $\Delta(\mathcal{R}', \mathcal{C}) \leq \Delta(\mathcal{R}, \mathcal{C}) - \sum_{i=1}^\alpha (k_i - 1)$ .*

*Proof.* Our proof is by induction. By Lemma 6, the lemma is true when  $\alpha = 1$ . So we will assume that the lemma is true for  $\alpha = 1, 2, \dots, l - 1$  where  $l \leq 2$ , and that the lemma has to be established for  $\alpha = l$ .

In the initial part of our proof (Stages 1 and 2 below) we produce another 2-partition  $\mathcal{R}^* = \{R_1^*, R_2^*, \dots, R_m^*\}$  satisfying a few conditions. Setting  $k_i^* = \Delta(R_i^*, \mathcal{C})$  for each  $i \in [m]$ , three of these conditions are  $k_l^* \geq 2$ ,  $\sum_{i=1}^l k_i^* \leq \sum_{i=1}^l k_i$ , and  $R_i^* = R_i$  for  $i > l$ . Note that this third condition implies

$$\Delta(\mathcal{R}, \mathcal{C}) - \sum_{i=1}^l (k_i - 1) = \sum_{i=1}^m k_i - \sum_{i=1}^l (k_i - 1) \tag{2}$$

$$= l + \sum_{i=l+1}^m k_i = l + \sum_{i=l+1}^m k_i^* = \Delta(\mathcal{R}^*, \mathcal{C}) - \sum_{i=1}^l (k_i^* - 1). \tag{3}$$

Initially, set  $\mathcal{R}^* = \mathcal{R}$ .

**Stage 1.** We recall that  $l \geq 2$ . From the conditions of the lemma (or the manner in which  $\mathcal{R}^*$  is recreated in Stage 2) either  $(k_l^* \geq 2$  and  $k_t^* \geq 1$  for some  $t \neq l)$  or  $k_l^* \geq 3$ .

Therefore, by Lemma 6, we can create a new row partition  $\mathcal{R}''$  from  $\mathcal{R}^*$  by performing a trade on  $\{R_l^*, R_x^*\}$  (for some  $x \neq l$ ), so that

$$\Delta(\mathcal{R}'', \mathcal{C}) \leq \Delta(\mathcal{R}^*, \mathcal{C}) - (k_l^* - 1). \quad (4)$$

Let  $\mathcal{R}'' = \{R_1'', R_2'', \dots, R_m''\}$ . Here, we let  $R_l''$  and  $R_x''$  be the row pairs created from  $R_l^*$  and  $R_x^*$  respectively after performing the trade, and let  $R_i'' = R_i^*$  for each  $i \notin \{l, x\}$ . Let  $k_i'' = \Delta(R_i'')$  for each  $i \in [m]$ . It follows that

$$\Delta(\mathcal{R}'', \mathcal{C}) - k_x'' - k_l'' = \Delta(\mathcal{R}^*, \mathcal{C}) - k_x^* - k_l^*. \quad (5)$$

The cases  $x > l$  and  $x < l$  have to be considered separately. If  $x < l$ , move on to Stage 2; if  $x > l$ , move on to Stage 3.

**Stage 2:**  $x < l$ .

In this case, provided either  $(k_i'' \geq 2, k_j'' \geq 1)$  for some  $i, j \leq l$ , or  $k_i'' \geq 3$  for some  $i \leq l$ , we re-order the row pairs  $R_i''$  for  $i \leq l$  so that  $k_i'' \geq 2$ . Now set  $R^* = R''$  and go back to the start of Stage 1, repeating the process in that stage with this new  $R^*$ . Note that  $k_l^* \geq 2$  for this new  $R^*$ , and, if  $k_l^* = 2$ , then  $k_j^* \geq 1$  for some  $j < l$ .

Eventually one of two things will happen: Either the required trade to obtain  $\mathcal{R}''$  from  $\mathcal{R}^*$  in Stage 1 will have  $x > l$  (in which case we move on to Stage 3) or, by reasoning similar to that in Corollary 7, we will reach a 2-partition  $\mathcal{R}''$  (here in Stage 2) such that either [1]  $k_i'' \in \{0, 1\}$  for all  $i \in [l]$  or [2]  $\sum_{i=1}^l k_i'' = 2$ . For  $l \geq 2$ , if either of these conditions on  $\mathcal{R}''$  holds, the total number of discrepancies in  $R_1, R_2, \dots, R_l$  will have been reduced by at least  $\sum_{i=1}^l (k_i - 1)$ , when going from  $\mathcal{R}$  to  $\mathcal{R}''$ , while  $R_i'' = R_i$  for  $l + 1 \leq i \leq m$ . Hence in the statement of the lemma we can take  $\mathcal{R}' = \mathcal{R}''$ .

**Stage 3:**  $x > l$ .

Here,  $\mathcal{R}^*$  and  $\mathcal{R}''$  are 2-partitions obtained after going through Stage 1 and obtaining  $x > l$ . Note that since  $\mathcal{R}''$  was obtained from  $\mathcal{R}^*$  by means of a trade between  $R_l^*$  and  $R_x^*$  for some  $x > l$ ,  $\{R_1^*, R_2^*, \dots, R_{l-1}^*\} \subset R''$ . So by applying our inductive hypothesis (Lemma 8 with  $\alpha = l - 1$ ) to  $\mathcal{R}''$ , unless  $k_1^* \leq 2$  and  $l = 2$  (dealt with below; here neither  $l - 1 \geq 2$  nor  $k_{l-1}^* \geq 3$ ) there exists a 2-partition  $\mathcal{R}'$  of the rows such that

$$\Delta(\mathcal{R}', \mathcal{C}) \leq \Delta(\mathcal{R}'', \mathcal{C}) - \sum_{i=1}^{l-1} (k_i^* - 1).$$

Combining this with (4),  $\Delta(\mathcal{R}', \mathcal{C}) \leq \Delta(\mathcal{R}^*, \mathcal{C}) - \sum_{i=1}^l (k_i^* - 1) = \Delta(\mathcal{R}, \mathcal{C}) - \sum_{i=1}^l (k_i - 1)$  as required.

So suppose  $l = 2$ , and  $k_1^* \leq 2$ . If  $k_1^* \leq 1$ , then  $k_1^* - 1 \leq 0$ , and so  $\Delta(\mathcal{R}'', \mathcal{C}) \leq \Delta(\mathcal{R}'', \mathcal{C}) - \sum_{i=1}^{l-1} (k_i^* - 1)$ . Combining this with (4) as above, gives  $\Delta(\mathcal{R}'', \mathcal{C}) \leq \Delta(\mathcal{R}^*, \mathcal{C}) - \sum_{i=1}^l (k_i^* - 1) = \Delta(\mathcal{R}, \mathcal{C}) - \sum_{i=1}^l (k_i - 1)$  as required.

Finally, suppose  $l = 2$  and  $k_1^* = 2$ . Here, from (4), we already have  $\Delta(\mathcal{R}'', \mathcal{C}) \leq \Delta(\mathcal{R}^*, \mathcal{C}) - (k_l^* - 1)$ . If either  $k_2'' \geq 1$  or  $k_x'' \geq 1$ , then (since  $k_1'' = k_1^* = 2$ ), we now have two discrepancies in one pair of rows and at least one in another; thus we may apply Lemma 6 to further reduce the number of discrepancies by 1, as needed. Otherwise  $k_2'' = k_x'' = 0$ . By (5) and the fact that  $k_x^* \geq 0$ ,

$$\begin{aligned}
\Delta(\mathcal{R}'', \mathcal{C}) &= \Delta(\mathcal{R}'', \mathcal{C}) - k_2'' - k_x'' = \Delta(\mathcal{R}^*, \mathcal{C}) - k_2^* - k_x^* \\
&= \Delta(\mathcal{R}^*, \mathcal{C}) - (k_1^* - 1) - (k_2^* - 1) + (k_1^* - 2 - k_x^*) \\
&\leq \Delta(\mathcal{R}^*, \mathcal{C}) - (k_1^* - 1) - (k_2^* - 1) \\
&= \Delta(\mathcal{R}, \mathcal{C}) - (k_1 - 1) - (k_2 - 1),
\end{aligned}$$

as required. □

**Theorem 9** *For any Latin square of order  $2m$  with  $m \geq 2$ , there exists  $k \in [m]$  and a pair of 2-partitions  $(\mathcal{R}, \mathcal{C})$  such that either  $\Delta(\mathcal{R}, \mathcal{C}) = 2$  or  $(k \leq (m + 3)/2, \Delta(\mathcal{R}, \mathcal{C}) = k$  and for each  $a \in [k]$ ,  $\Delta(R_a) = \Delta(C_a) = 1$  and  $S(R_a, C_a)$  contains a discrepancy).*

*Proof.* Assume that all conditions of the lemma hold, except that we have  $k > (m + 3)/2$ . We will show (by contradiction, supposing the opposite) that it is possible to create 2-partitions of the rows and columns which decrease  $\Delta$ . By then repeatedly applying Lemma 6 (if necessary), the theorem holds. We focus on the potentials within column pairs  $C_1 \cup C_2 \cup \dots \cup C_k$ .

Suppose first that there exist  $\{i, j\} \subset [k]$  and  $t \in \{1, 2\}$  such that there is at most one potential of type  $t$  of the form  $(s, R_l, C_i, C_j)$  with  $l > k$ , and that for any potential of type  $t$  of the form  $(s, R_l, C_i, C_j)$ , we have  $l \notin \{i, j\}$ .

Let  $\mathcal{C}'$  be the 2-partition of the columns after making a trade of type  $t$  on  $\{C_i, C_j\}$  for such  $i, j, t$  as described above. From the conditions of the theorem,  $d^-(C_i \leftrightarrow C_j, t) = 2$ . Let  $d^+(C_i \leftrightarrow C_j, t) = \beta$ . Then,  $\Delta(\mathcal{R}, \mathcal{C}') = \Delta(\mathcal{R}, \mathcal{C}) + \beta - 2$ . Thus if  $\beta \leq 1$  we are done. Otherwise, if we let  $C'_i, C'_j$  be the two new pairs of columns after the trade, then at most one of the  $\beta$  new discrepancies created is not of the form  $(s, R_z, C'_i, C'_j)$  with  $z \in [k]$ ,  $z \notin \{i, j\}$ . Thus at most one new discrepancy is not in a pair of rows containing a discrepancy that existed both before and after the trade. In this case, reorder the row pairs  $(R_y : y \leq k)$  by interchanging  $R_i$  with  $R_{k-1}$ ,  $R_j$  with  $R_k$  and  $R_{k-2}$  with  $R_x$  for some other  $x \in [k - 2]$  such that  $\Delta(R_x, \mathcal{C}') \geq 2$ . We can now apply Lemma 8 (with  $\alpha = k - 2$ ) to produce another 2-partition of the rows and columns with at least  $\beta - 1$  fewer discrepancies. Overall we have reduced the number of discrepancies by at least  $\beta - 1 \geq 1$ , so we are done.

Otherwise, for each  $\{i, j\} \subset [k]$  and each  $t \in \{1, 2\}$ , we must have either:

1. at least one potential of type  $t$  of the form  $(s, R_z, C_i, C_j)$  with  $z \in \{i, j\}$ , or
2. at least two potentials of type  $t$  of the form  $(s, R_z, C_i, C_j)$ , with  $z > k$ .

We consider these potentials to be of *variety* 1 and 2, respectively. Let  $v_x$  be the total number of occurrences of variety  $v_x$ , for each  $x \in [2]$ . We proceed to obtain some inequalities involving these parameters. Each occurrence of variety 1 uses at least one symbol from a subarray  $S(R_z, C_z)$  with  $z \in [k]$ ; moreover each such subarray  $S(R_z, C_z)$  already contains a discrepancy; thus  $2k \geq v_1$ . Next, there are  $4(m - k)k$  cells within both the first  $2k$  columns and the final  $2(m - k)$  rows; thus  $4(m - k)k \geq 4v_2$ , i.e.  $(m - k)k \geq v_2$ .

Finally, there are  $k(k-1)$  choices for  $i, j$  and type  $t$ , so we must have  $k(k-1) \leq v_1 + v_2$ . But combining the above two inequalities for  $v_1, v_2$ , gives  $v_1 + v_2 \leq 2k + (m-k)k$ . Therefore,  $k(k-1) \leq 2k + (m-k)k$ , or equivalently,  $2k^2 \leq (m+3)k$ . It follows that  $k \leq (m+3)/2$ , a contradiction.  $\square$

## 4 Motivation

The original motivation for studying this problem is related to the study of defining sets and trades in Latin squares. If a subarray of a Latin square contains no repeated symbols, it is not hard to see that the complement of the subarray within the Latin square has a unique completion. Equivalently, if a subarray has no repeated symbols then it contains no Latin trades. See [1] for relevant definitions and a survey of these concepts. Thus the existence of subarrays without repeated symbols may have implications for the study of defining sets and Latin trades in Latin squares.

However, independent of possible application, the problem studied in this paper seems to be a natural enough theoretical question in its own right.

## References

- [1] N. J. Cavenagh, *The theory and application of latin bitrades: a survey*, Math. Slovaca 58 (2008), 1–28.
- [2] C. J. Colbourn, J. H. Dinitz and I. M. Wanless, Latin Squares, in: *The CRC Handbook of Combinatorial Designs, 2nd ed.* (C. J. Colbourn and J. H. Dinitz, eds.), CRC Press, Boca Raton, FL, (2007), 135–152.
- [3] B. D. McKay, Latin squares, <http://cs.anu.edu.au/~bdm/data/latin.html>.